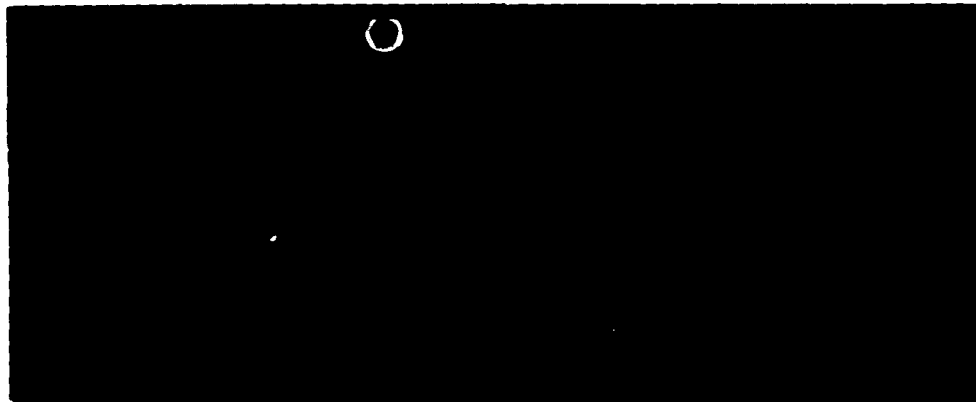


AD-A219 256

DTIC FILE COPY

①



# PURDUE UNIVERSITY



**DTIC**  
**ELECTE**  
**MAR 15 1990**  
**S E D**

**CENTER FOR STATISTICAL  
DECISION SCIENCES AND  
DEPARTMENT OF STATISTICS**

**DISTRIBUTION STATEMENT A**

Approved for public release;  
Distribution Unlimited

90 03 14 083

SELECTING MULTINOMIAL POPULATIONS\*

by

Shanti S. Gupta  
Department of Statistics  
Purdue University  
West Lafayette, IN 47907

and TaChen Liang  
Department of Mathematics  
Wayne State University  
Detroit, MI 48202

Technical Report #90-01C

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input checked="" type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Department of Statistics  
Purdue University

January 1990

DTIC  
ELECTE  
MAR 15 1990

\* This research was supported in part by the Office of Naval Research Contract N00014-88-K-0170 and NSF Grants DMS-8606964, DMS-8702620, and DMS-8717799

# Selecting Multinomial Populations\*

by

Shanti S. Gupta  
Department of Statistics  
Purdue University  
West Lafayette, IN 47907

TaChen Liang  
Department of Mathematics  
Wayne State University  
Detroit, MI 48202

## ABSTRACT

This paper deals with the problem of selecting fair multinomial populations compared with a standard. Two selection procedures are investigated: the natural selection procedure of Gupta and Leu (1989) and an empirical Bayes simultaneous selection procedure. It is proved that the natural selection procedure is a Bayes procedure relative to a symmetric Dirichlet prior distribution, and therefore is an admissible selection procedure. For the empirical Bayes simultaneous selection procedure, the associated asymptotic optimality is investigated. It is shown that the proposed empirical Bayes selection procedure is asymptotically optimal relative to a class of symmetric Dirichlet priors. The rate of convergence of the empirical Bayes selection procedure is shown to be of order  $O(\exp(-\tau k + \ln k))$  for some positive constant  $\tau$ , where  $k$  is the number of populations involved in the selection problem.

AMS Classification: Primary 62F07; secondary 62C12.

Key Words and Phrases: fair multinomial population; entropy function; Gini-Simpson index; natural selection procedure; empirical Bayes selection procedure; asymptotically optimal; rate of convergence.

---

\* This research was supported in part by the Office of Naval Research Contract N00014-88-K-0170 and NSF Grants DMS-8606964, DMS-8702620, and DMS-8717799

## 1. INTRODUCTION

The concept of diversity within a population is of considerable importance in statistical theory and applications. The problem of measuring diversity arises in a variety of studies in ecology, sociology, econometrics, genetics and many other sciences. For a multinomial population with  $m$  cells, the index of diversity is a function of the corresponding probability parameter vector  $\underline{p} = (p_1, \dots, p_m)$ . In practice, a Schur-convex or Schur-concave function of  $\underline{p}$  may be appropriate. There are two measures of diversity of a multinomial population which have been commonly used. These are the Shannon's entropy function and the Gini-Simpson index. The notion of the entropy function was introduced by Shannon (1948). The Gini-Simpson index was introduced by Gini (1912) and Simpson (1949). Both are Schur-concave function of  $\underline{p}$ .

In the literature, selection procedures using indices of diversity as selection criteria have been studied by many authors. Gupta and Huang (1976) studied the problem of selecting the population with the largest entropy function for binomial distributions. Gupta and Wong (1975) considered the problem of selecting a subset containing the population with the largest entropy for multinomial distributions. Dudewicz and Van der Meulen (1981) investigated a selection procedure based on a generalized entropy function. Alam, Mitra, Rizvi and Saxena (1986) studied selection procedures based on the Shannon's entropy function and the Gini-Simpson index using the indifference zone approach. Rizvi, Alam and Saxena (1987) also considered a subset selection procedure based on certain other diversity indexes. Recently, Gupta and Leu (1989) have studied certain selection procedures based on the Gini-Simpson index.

In this paper, we are dealing with the problem of selecting fair populations compared with a standard level. Consider  $k$  independent multinomial populations  $\pi_1, \dots, \pi_k$ . For each  $i = 1, \dots, k$ , population  $\pi_i$  has  $m$  cells and is characterized by the corresponding probability parameter vector  $\underline{p}_i = (p_{i1}, \dots, p_{im})$ , where  $0 \leq p_{ij} \leq 1$ ,  $j = 1, \dots, m$ , and  $\sum_{j=1}^m p_{ij} = 1$  for each  $i = 1, \dots, k$ . Define

$$\theta_i = \Psi(\underline{p}_i) = \sum_{j=1}^m \left( p_{ij} - \frac{1}{m} \right)^2. \quad (1.1)$$

We use  $\theta_i$  as a measure of diversity (or uniformity) of population  $\pi_i$ . Note that since

$\theta_i = \sum_{j=1}^m p_{ij}^2 - \frac{1}{m}$ , it is essentially equivalent to the Gini-Simpson index. Also note that  $0 \leq \theta_i \leq 1 - \frac{1}{m}$ . For a given constant  $\theta_0$ ,  $0 < \theta_0 < 1 - \frac{1}{m}$ , population  $\pi_i$  is said to be a fair population if  $\theta_i \leq \theta_0$  and a bad population, otherwise. Our goal is to derive statistical selection procedures for selecting all fair populations while excluding all bad populations. It should be noted that the problem of selecting fair multinomial populations has been considered by Gupta and Leu (1989) through a classical approach.

Let  $\Omega = \{\underline{p} = (p_1, \dots, p_k) | \underline{p}_i = (p_{i1}, \dots, p_{im}), 0 \leq p_{ij} \leq 1, j = 1, \dots, m, \sum_{j=1}^m p_{ij} = 1 \text{ for each } i = 1, \dots, k\}$  be the parameter space and let  $\mathcal{A} = \{s | s \subset \{1, \dots, k\}\}$  be the action space. When action  $s$  is taken, it means that population  $\pi_i$  is selected as a fair population if  $i \in s$  and excluded as a bad population if  $i \notin s$ . For  $\underline{p} \in \Omega$  and action  $s \in \mathcal{A}$ , the loss function  $L(\underline{p}, s)$  is defined to be:

$$L(\underline{p}, s) = \sum_{i \notin s} (\theta_i - \theta_0) I_{(\theta_0, 1 - \frac{1}{m}]}(\theta_i) + \sum_{i \in s} (\theta_0 - \theta_i) I_{[0, \theta_0]}(\theta_i). \quad (1.2)$$

In (1.2), the first summation is the loss due to selecting certain bad populations and the second summation is the loss due to not selecting certain fair populations.

The content of this paper consists of two parts. In Section 2, we investigate some optimal properties of the natural selection procedure of Gupta and Leu (1989). It is shown that, for the loss function  $L(\underline{p}, s)$  of (1.2), the natural selection procedure is Bayes relative to some symmetric Dirichlet prior, and therefore, it is admissible. Section 3 deals with this selection problem through a parametric empirical Bayes approach. An empirical Bayes selection procedure is proposed and the corresponding asymptotic optimality is investigated. It is shown that the proposed empirical Bayes selection procedure is asymptotically optimal relative to a class of symmetric Dirichlet priors. The rate of convergence of the proposed empirical Bayes selection procedure is also established, and shown to be of order  $O(\exp(-\tau k + \ln k))$  for some positive constant  $\tau$ , where  $k$  is the number of populations involved in the selection problem.

## 2. OPTIMALITY OF NATURAL SELECTION PROCEDURE

For each  $i = 1, \dots, k$ , let  $\underline{X}_i = (X_{i1}, \dots, X_{im})$  be the random observation associated with population  $\pi_i$ , where  $X_{ij}$ ,  $1 \leq j \leq m$ , are nonnegative integer random variables such

that  $0 \leq X_{ij} \leq N$  and  $\sum_{j=1}^m X_{ij} = N$ . Then,  $\underline{X}_i$  has the probability function

$$f(\underline{x}_i | \underline{p}_i) = \frac{N!}{\prod_{j=1}^m (x_{ij}!)} \prod_{j=1}^m p_{ij}^{x_{ij}-1} \quad (2.1)$$

at point  $\underline{x}_i = (x_{i1}, \dots, x_{im})$  for which  $0 \leq x_{ij} \leq N$ ,  $1 \leq j \leq m$  and  $\sum_{j=1}^m x_{ij} = N$ . Let  $\mathcal{X}_i$  be the sample space generated by  $\underline{X}_i$ . Let  $\underline{X} = (\underline{X}_1, \dots, \underline{X}_k)$  and denote the corresponding observed value by  $\underline{x} = (\underline{x}_1, \dots, \underline{x}_k)$ . Also, let  $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_k$  denote the sample space of  $\underline{X}$ .

A selection procedure  $d = (d_1, \dots, d_k)$  is defined to be a mapping from the sample space  $\mathcal{X}$  into the product space  $[0, 1]^k$ . That is, for each  $\underline{x} = (\underline{x}_1, \dots, \underline{x}_k) \in \mathcal{X}$ ,  $d(\underline{x}) = (d_1(\underline{x}), \dots, d_k(\underline{x}))$ , where  $d_i(\underline{x})$  is the probability of selecting population  $\pi_i$  as a fair population given  $\underline{X} = \underline{x}$  is observed. We let  $D$  denote the class of all selection procedures defined in the above way.

For each  $i = 1, \dots, k$ ,  $(\frac{X_{i1}}{N}, \dots, \frac{X_{im}}{N})$  is the maximum likelihood estimator of  $\underline{p}_i = (p_{i1}, \dots, p_{im})$ . From (1.1), it is natural and reasonable to estimate  $\theta_i$  by  $\hat{\theta}_i = \sum_{j=1}^m (\frac{X_{ij}}{N})^2 - \frac{1}{m}$ . Gupta and Leu (1989) proposed a natural selection procedure  $d^N = (d_1^N, \dots, d_k^N)$  based on  $\hat{\theta}_i$ ,  $i = 1, \dots, k$ , which is equivalent to the following: For each  $i = 1, \dots, k$ ,

$$d_i^N(\underline{X}) = \begin{cases} 1 & \text{if } \hat{\theta}_i \leq \delta, \\ 0 & \text{otherwise,} \end{cases} \quad (2.2)$$

where  $\delta$  is a prespecified positive constant such that  $0 < \delta < 1 - \frac{1}{m}$ . Since this natural selection procedure is heavily dependent on the constant  $\delta$ , we denote this procedure by  $d^{N(\delta)} = (d_1^{N(\delta)}, \dots, d_k^{N(\delta)})$ .

In the following, it is assumed that for each  $i = 1, \dots, k$ , the parameter vector  $\underline{p}_i = (p_{i1}, \dots, p_{im})$  is a realization of the random vector  $\underline{P}_i = (P_{i1}, \dots, P_{im})$ . It is also assumed that  $\underline{P}_1, \dots, \underline{P}_k$  are iid with a common prior distribution  $G_\alpha$  belonging to a class of symmetric Dirichlet distributions  $\mathcal{C}$ , where

$$\mathcal{C} = \{g_\alpha | g_\alpha(\underline{p}_i) = \frac{\Gamma(m\alpha)}{[\Gamma(\alpha)]^m} \prod_{j=1}^m p_{ij}^{\alpha-1}, \quad 0 \leq p_{ij} \leq 1, \quad j = 1, \dots, m, \quad \sum_{j=1}^m p_{ij} = 1\}. \quad (2.3)$$

For a prior distribution  $G_\alpha \in \mathcal{C}$  and a selection procedure  $d = (d_1, \dots, d_k) \in D$ , we denote the corresponding Bayes risk by  $r(G_\alpha, d)$ . From (1.2) and the statistical model described previously,

$$r(G_\alpha, d_i) = \sum_{i=1}^k r_i(G_\alpha, d_i) \quad (2.4)$$

where

$$\begin{aligned} r_i(G_\alpha, d_i) = & \sum_{\underline{x} \in \mathcal{X}} d_i(\underline{x}) \left[ \sum_{j=1}^m E[P_{ij}^2 | \underline{x}_i] - \theta_0 \right] \prod_{j=1}^k f(\underline{x}_j) \\ & + \sum_{\underline{x} \in \mathcal{X}} \int_{\Omega_i(\theta_0)} (\theta_0 - \theta_i) \prod_{j=1}^k [f(\underline{x}_j | \underline{p}_j) g_\alpha(\underline{p}_j)] d\underline{p}, \end{aligned} \quad (2.5)$$

$$f(\underline{x}_i) = \int f(\underline{x}_i | \underline{p}_i) g_\alpha(\underline{p}_i) d\underline{p}_i = \frac{N!}{\prod_{j=1}^m (x_{ij}!)} \frac{\Gamma(m\alpha)}{[\Gamma(\alpha)]^m} \frac{\prod_{j=1}^m \Gamma(\alpha + x_{ij})}{\Gamma(m\alpha + N)},$$

and  $\Omega_i(\theta_0) = \{\underline{p} = (\underline{p}_1, \dots, \underline{p}_k) \in \Omega | \theta_i \leq \theta_0\}$ .

Since the second term of (2.5) does not depend on the selection procedure  $d$ , a Bayes selection procedure  $d^{G_\alpha} = (d_1^{G_\alpha}, \dots, d_k^{G_\alpha})$  can be obtained as follows: For each  $i = 1, \dots, k$ ,

$$d_i^{G_\alpha}(\underline{x}) = \begin{cases} 1 & \text{if } \sum_{j=1}^m E[P_{ij}^2 | \underline{x}_i] - \frac{1}{m} \leq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \quad (2.6)$$

Then, we have the following theorem.

**Theorem 2.1.** For each positive constant  $\delta$  such that  $0 < \delta < 1 - \frac{1}{m}$  and  $\frac{m\delta N^2 + mN - N}{mN(N+1)} > \theta_0$ , for the loss function  $L(\underline{p}, s)$ , the natural selection procedure  $d^{N(\delta)}$  given in (2.2) is a Bayes procedure relative to some symmetric prior distribution  $G_\alpha$ .

**Proof:** First, straightforward computation yields that for each  $i = 1, \dots, k$  and  $j = 1, \dots, m$ ,

$$E[P_{ij}^2 | \underline{x}_i] = \frac{(x_{ij} + \alpha + 1)(x_{ij} + \alpha)}{(m\alpha + N + 1)(m\alpha + N)}$$

and therefore

$$\sum_{j=1}^m E[P_{ij}^2 | \underline{x}_i] - \frac{1}{m} = \frac{(m^2 - m)\alpha + mN + m \sum_{j=1}^m x_{ij}^2 - N^2 - N}{m(m\alpha + N + 1)(m\alpha + N)}. \quad (2.7)$$

Note that

$$\hat{\theta}_i \leq \delta \iff \sum_{j=1}^m E[P_{ij}^2 | \underline{x}_i] - \frac{1}{m} \leq H(\alpha)$$

where

$$H(\alpha) = \frac{(\delta + \frac{1}{m})mN^2 + (m^2 - m)\alpha + mN - N^2 - N}{m(m\alpha + N + 1)(m\alpha + N)}.$$

Thus, it suffices to prove that for given  $\theta_0$ ,  $0 < \theta_0 < 1 - \frac{1}{m}$ , there exists a positive  $\alpha$  such that  $H(\alpha) = \theta_0$ . Since  $H(\alpha)$  is decreasing in  $\alpha$ ,  $H(0) = \frac{m\delta N^2 + mN - N}{mN(N+1)} > \theta_0$  by the assumption,  $\lim_{\alpha \rightarrow \infty} H(\alpha) = 0 < \theta_0$  and  $H(\alpha)$  is a continuous function of  $\alpha$  on  $[0, \infty)$ , there exists a unique  $\alpha \equiv \alpha(\theta_0) > 0$  such that  $H(\alpha) = \theta_0$ . This implies that the natural selection procedure  $d^{N(\delta)}$  is the Bayes procedure relative to the symmetric Dirichlet prior distribution  $G_{\alpha(\theta_0)}$ . Hence the proof of this theorem is complete.  $\square$

The following corollary is a direct consequent of Theorem 2.1.

**Corollary 2.1.** For each positive constant  $\delta$  such that  $0 < \delta < 1 - \frac{1}{m}$ ,  $\frac{m\delta N^2 + mN - N}{mN(N+1)} > \theta_0$ , the natural selection procedure is admissible for the loss function  $L(\underline{p}, s)$ .

### 3. AN EMPIRICAL BAYES SELECTION PROCEDURE

We assume that the hyperparameter  $\alpha$  of the symmetric Dirichlet prior distribution  $G_\alpha$  is unknown. In this situation, it is not possible to apply the Bayes selection procedure  $d^{G_\alpha}$  for the selection problem at hand. Thus, the empirical Bayes approach is employed here.

For each  $i = 1, \dots, k$ , let  $W_i = \sum_{j=1}^m X_{ij}^2$  and let  $w_i$  denote the observed value of  $W_i$ . Under the preceding statistical model,  $W_1, \dots, W_k$  are iid random variables such that  $\frac{N^2}{m} \leq W_i \leq N^2$ . It follows from straightforward computations that

$$\mu_2 = E[W_1] = N \left[ 1 + \frac{(N-1)(\alpha+1)}{m\alpha+1} \right], \quad (3.1)$$

and therefore,

$$\alpha = \frac{N^2 - \mu_2}{m\mu_2 - N(m+N-1)}. \quad (3.2)$$



From (3.2) and (2.7), for each  $i = 1, \dots, k$ ,

$$\begin{aligned} & \sum_{j=1}^m E[P_{ij}^2 | \underline{x}_i] - \frac{1}{m} \\ &= \{(m^2 - m)(N^2 - \mu_2) + [m\mu_2 - N(m + N - 1)][mN + mw_i - N^2 - N]\} \\ & \quad \times (m\mu_2 - mN - N^2 + N) / [m(N - 1)(m\mu_2 - N^2)N(m\mu_2 - N^2 - m + 1)] \\ & \equiv Q_i(\mu_2 | w_i), \end{aligned} \quad (3.3)$$

where  $w_i = \sum_{j=1}^m x_{ij}^2$ .

Note that for  $w_i$  (or  $\underline{x}_i$ ) being kept fixed,  $Q_i(\mu_2 | w_i)$  is increasing in  $\mu_2$ .

Define,  $\hat{\mu}_2 = \frac{1}{k} \sum_{i=1}^k W_i$ . We will use  $\hat{\mu}_2$  to estimate  $\mu_2$  and use  $Q_i(\hat{\mu}_2 | w_i)$  to estimate  $Q_i(\mu_2 | w_i)$ . However, by the definition of  $\mu_2$  (see (3.1)),  $N + \frac{N(N-1)}{m} < \mu_2 < N^2$  and  $Q_i(\mu_2 | w_i)$  tends to zero as  $\mu_2$  tends to  $N + \frac{N(N-1)}{m}$ . Also, it is possible that  $\hat{\mu}_2 \leq N + \frac{N(N-1)}{m}$ . Thus we define

$$\varphi_i^*(w_i) = \begin{cases} Q_i(\hat{\mu}_2 | w_i) & \text{if } \hat{\mu}_2 > N + \frac{N(N-1)}{m}, \\ 0 & \text{otherwise.} \end{cases} \quad (3.4)$$

We now propose an empirical Bayes selection procedure  $d^* = (d_1^*, \dots, d_k^*)$  as follows: For each  $i = 1, \dots, k$ ,

$$d_i^*(\underline{x}) = \begin{cases} 1 & \text{if } \varphi_i^*(w_i) \leq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \quad (3.5)$$

In the following, we will investigate the asymptotic performance of the empirical Bayes selection procedure  $d^*$  for the case where  $k$ , the number of populations involved in the selection problem under study, is sufficiently large.

Since  $d^{G_\alpha} = (d_1^{G_\alpha}, \dots, d_k^{G_\alpha})$  is the Bayes selection procedure, for the empirical Bayes selection procedure  $d^* = (d_1^*, \dots, d_k^*)$ ,  $r_i(G_\alpha, d_i^*) - r_i(G_\alpha, d_i^{G_\alpha}) \geq 0$  for each  $i = 1, \dots, k$  and therefore,  $r(G_\alpha, d^*) - r(G_\alpha, d^{G_\alpha}) = \sum_{i=1}^k [r_i(G_\alpha, d_i^*) - r_i(G_\alpha, d_i^{G_\alpha})] \geq 0$ . This nonnegative regret value  $r(G_\alpha, d^*) - r(G_\alpha, d^{G_\alpha})$  will be used as a measure of performance of the empirical Bayes selection procedure  $d^*$ .

**Definition 3.1.** A selection procedure  $d = (d_1, \dots, d_k) \in D$  is said to be asymptotically optimal of order  $\{\beta_k\}$  relative to the prior distribution  $G$  if

$$r(G, d^*) - r(G, d^G) = O(\beta_k)$$

where  $\{\beta_k\}$  is a sequence of positive numbers such that  $\lim_{k \rightarrow \infty} \beta_k = 0$ .

For each  $i = 1, \dots, k$  and for the fixed  $\mu_2$ ,  $Q_i(\mu_2|w_i)$ , which is defined in (3.3), can be viewed as a function of  $w_i$ . It is clear that  $Q_i(\mu_2|w_i)$  is increasing in  $w_i$ . Let

$$A_i = \{w_i | w_i = w_i(\underline{x}_i) = \sum_{j=1}^m x_{ij}^2, \underline{x}_i \in \mathcal{X}_i, Q_i(\mu_2|w_i) < \theta_0\}$$

$$B_i = \{w_i | w_i = w_i(\underline{x}_i) = \sum_{j=1}^m x_{ij}^2, \underline{x}_i \in \mathcal{X}_i, Q_i(\mu_2|w_i) > \theta_0\}$$

From the statistical model under consideration,  $Q_i(\mu_2|w) = Q_j(\mu_2|w) = Q(\mu_2|w)$  (say) for all  $i, j = 1, \dots, k$ . Thus,  $A_1 = \dots = A_k = A$  and  $B_1 = \dots = B_k = B$  (say).

Let  $h(w)$  denote the common marginal probability function of the iid random variables  $W_i = \sum_{j=1}^m X_{ij}^2$ ,  $i = 1, \dots, k$ . From (2.5)–(2.6) and (3.5), straightforward computation yields that

$$\begin{aligned} 0 &\leq r_i(G_\alpha, d_i^*) - r_i(G_\alpha, d_i^{G_\alpha}) \\ &\leq \sum_{w_i \in A} P_i\{\varphi_i^*(w_i) > \theta_0 | W_i = w_i\} h(w_i) \\ &\quad + \sum_{w_i \in B} P_i\{\varphi_i^*(w_i) \leq \theta_0 | W_i = w_i\} h(w_i) \end{aligned} \quad (3.6)$$

where the probability measure  $P_i$  is computed with respect to  $(W_1, \dots, W_{i-1}, W_{i+1}, \dots, W_k)$ .

Thus, it suffices to investigate the asymptotic behavior of  $P_i\{\varphi_i^*(w_i) \leq \theta_0 | W_i = w_i\}$  for  $w_i \in B$  and  $P_i\{\varphi_i^*(w_i) > \theta_0 | W_i = w_i\}$  for  $w_i \in A$ .

**Lemma 3.1.** For each  $c > 0$  and for sufficiently large  $k$ ,

- (a)  $P_i\{\hat{\mu}_2 - \mu_2 < -c | W_i = w_i\} \leq O(\exp(-kc^2 N^{-4}(1 - \frac{1}{m})^{-2}/2))$ .
- (b)  $P_i\{\hat{\mu}_2 - \mu_2 > c | W_i = w_i\} \leq O(\exp(-kc^2 N^{-4}(1 - \frac{1}{m})^{-2}/2))$ .
- (c)  $P_i\{\hat{\mu}_2 \leq N + \frac{N(N-1)}{m} | W_i = w_i\} \leq O(\exp(-k(\mu_2 - N - (N-1)N/m)^2 N^{-4}(1 - \frac{1}{m})^{-2}/2))$ .

Note that the above upper bounds are independent of the value  $w_i$ .

**Proof:** (a). Let  $\hat{\mu}_2(i) = \frac{1}{k-1} \sum_{\substack{j=1 \\ j \neq i}}^k W_j$ . Then,

$$P_i\{\hat{\mu}_2 - \mu_2 < -c | W_i = w_i\} = P_i\{\hat{\mu}_2(i) - \mu_2 < -\frac{kc}{k-1} + \frac{\mu_2 - w_i}{k-1}\}.$$

Note that  $\frac{N^2}{m} \leq w_i \leq N^2$  for all  $w_i = w_i(\underline{x}_i) = \sum_{j=1}^m x_{ij}^2$ . Thus for  $k$  sufficiently large  $-\frac{kc}{k-1} + \frac{\mu_2 - w_i}{k-1} \leq -\frac{c}{2}$ . Hence, we obtain, for  $k$  sufficiently large, that

$$\begin{aligned} P_i\{\hat{\mu}_2 - \mu_2 < -c | W_i = w_i\} \\ \leq P_i\{\hat{\mu}_2(i) - \mu_2 < -\frac{c}{2}\} \\ \leq \exp\{-kc^2 N^{-4} (1 - \frac{1}{m})^{-2} / 2\}, \end{aligned}$$

where the last inequality follows from Theorem 2 of Hoeffding (1963). Note that the upper bound is independent of  $w_i$ .

The proof of part (b) is similar to that of part (a). By letting  $c = \mu_2 - [N + \frac{N(N-1)}{m}]$ , then  $c > 0$ . Thus, the result of part (c) follows directly from part (a).  $\square$

**Lemma 3.2.** For  $w_i \in A$ ,

$$P_i\{\varphi_i^*(w_i) > \theta_0 | W_i = w_i\} \leq O(\exp(-k(Q_i^{-1}(\theta_0 | w_i) - \mu_2)^2 N^{-4} (1 - \frac{1}{m})^{-2} / 2))$$

where  $Q_i^{-1}(\cdot | w_i)$  is the inverse function of  $Q_i(\cdot | w_i)$ .

**Proof:** From (3.4) and the fact that  $\theta_0 > 0$ ,

$$P\{\varphi_i^*(w_i) > \theta_0 | W_i = w_i\} = P\{Q_i(\hat{\mu}_2 | w_i) > \theta_0 | W_i = w_i\}. \quad (3.7)$$

Now, for each fixed  $w_i \in A$ ,  $Q_i(\mu | w_i)$  is strictly increasing in  $\mu$  for  $N + \frac{N(N-1)}{m} < \mu < N^2$  and  $Q_i(\mu_2 | w_i) < \theta_0$ . Thus  $\mu_2 < Q_i^{-1}(\theta_0 | w_i)$ . Then,

$$\begin{aligned} Q_i(\hat{\mu}_2 | w_i) > \theta_0 &\iff \hat{\mu}_2 > Q_i^{-1}(\theta_0 | w_i) \\ &\iff \hat{\mu}_2 - \mu_2 > Q_i^{-1}(\theta_0 | w_i) - \mu_2 > 0. \end{aligned} \quad (3.8)$$

Combining (3.7) and (3.8), by Lemma 3.1 (b), we obtain, for  $w_i \in A$ ,

$$\begin{aligned} P_i\{\varphi_i^*(w_i) > \theta_0 | W_i = w_i\} \\ &= P_i\{\hat{\mu}_2 - \mu_2 > Q_i^{-1}(\theta_0 | w_i) - \mu_2 | W_i = w_i\} \\ &\leq O(\exp(-k(Q_i^{-1}(\theta_0 | w_i) - \mu_2)^2 N^{-4}(1 - \frac{1}{m})^{-2}/2)). \end{aligned}$$

Thus, the proof of this lemma is complete.  $\square$

**Lemma 3.3.** For  $w_i \in B$ ,

$$\begin{aligned} P_i\{\varphi_i^*(w_i) \leq \theta_0 | W_i = w_i\} &\leq O(\exp(-k(Q_i^{-1}(\theta_0 | w_i) - \mu_2)^2 N^{-4}(1 - \frac{1}{m})^{-2}/2)) \\ &\quad + O(\exp(-k(\mu_2 - N - (N-1)N/m)^2 N^{-4}(1 - \frac{1}{m})^{-2}/2)). \end{aligned}$$

**Proof:**

$$\begin{aligned} P_i\{\varphi_i^*(w_i) \leq \theta_0 | W_i = w_i\} \\ &= P_i\{\varphi_i^*(w_i) \leq \theta_0, \hat{\mu}_2 \leq N + \frac{N(N-1)}{m} | W_i = w_i\} \\ &\quad + P_i\{\varphi_i^*(w_i) \leq \theta_0, \hat{\mu}_2 > N + \frac{N(N-1)}{m} | W_i = w_i\}. \end{aligned} \tag{3.9}$$

From Lemma 3.1 (c),

$$\begin{aligned} P_i\{\varphi_i^*(w_i) \leq \theta_0, \hat{\mu}_2 \leq N + \frac{N(N-1)}{m} | W_i = w_i\} \\ \leq O(\exp(-k(\mu_2 - N - (N-1)N/m)^2 N^{-4}(1 - \frac{1}{m})^{-2}/2)). \end{aligned} \tag{3.10}$$

From (3.4) and an argument analogous to that given in the proof of Lemma 3.2, we have

$$\begin{aligned} P_i\{\varphi_i^*(w_i) \leq \theta_0, \hat{\mu}_2 > N + \frac{N(N-1)}{m} | W_i = w_i\} \\ &= P_i\{Q_i(\hat{\mu}_2 | w_i) \leq \theta_0, \hat{\mu}_2 > N + \frac{N(N-1)}{m} | W_i = w_i\} \\ &\leq P_i\{\hat{\mu}_2 - \mu_2 \leq Q_i^{-1}(\theta_0 | w_i) - \mu_2 | W_i = w_i\} \\ &= O(\exp(-k(Q_i^{-1}(\theta_0 | w_i) - \mu_2)^2 N^{-4}(1 - \frac{1}{m})^{-2}/2)) \end{aligned} \tag{3.11}$$

where  $Q_i^{-1}(\theta_0|w_i) - \mu_2 < 0$  since  $w_i \in B$ . Thus, the lemma follows from (3.9)–(3.11)  $\square$

Let  $\tau_1 = \min\{(Q_i^{-1}(\theta_0|w_i) - \mu_2)^2 N^{-4} (1 - \frac{1}{m})^{-2} / 2 | w_i \in A_i \cup B_i\}$ . Then,  $\tau_1 > 0$  since  $Q_i^{-1}(\theta_0|w_i) - \mu_2 \neq 0$  for all  $w_i \in A_i \cup B_i$  and  $A_i \cup B_i$  is a finite set. Then  $\tau \equiv \min(\tau_1, (\mu_2 - N - (N - 1)N/m)^2 N^{-4} (1 - \frac{1}{m})^{-2} / 2) > 0$ .

The following theorem describes the asymptotic optimality property of the empirical Bayes selection procedure  $d^* = (d_1^*, \dots, d_k^*)$ .

**Theorem 3.1.** Let  $d^* = (d_1^*, \dots, d_k^*)$  be the empirical Bayes selection procedure defined through (3.4)–(3.5). Suppose that the prior distribution is a member of the class  $\mathcal{C}$  of symmetric Dirichlet distributions given in (2.3). Then

(a) For each  $i = 1, \dots, k$ ,  $r_i(G_\alpha, d_i^*) - r_i(G_\alpha, d_i^{G_\alpha}) \leq O(\exp(-\tau k))$ , and

(b)  $r(G_\alpha, d^*) - r(G_\alpha, d^{G_\alpha}) \leq O(\exp(-\tau k + \ln k))$

where  $\tau$  is the positive constant defined previously.

**Proof:** Part (b) is a direct result of part (a). Thus, we need to prove part (a) only. From (3.6) Lemmas 3.2 and 3.3,

$$\begin{aligned} & r_i(G_\alpha, d_i^*) - r_i(G_\alpha, d_i^{G_\alpha}) \\ & \leq O(\exp(-\tau k)) \sum_{w_i \in A \cup B} h(w_i) \\ & = O(\exp(-\tau k)). \end{aligned}$$

Thus, the theorem follows.  $\square$

## REFERENCES

- Alam, K., Mitra, A., Rizvi, M. H. and Saxena, K. M. L. (1986). Selection of the most diverse multinomial population. *American Journal of Mathematical and Management Sciences*, Vol. 6, 65-86.
- Dudewicz, E. J. and Van der Meulen, E. C. (1981). Selection procedures for the best binomial population with generalized entropy goodness. *Tamkang Journal of Mathematics*, 12, 206-208.
- Gini, C. (1912). Variabilità e Mutabilità. *Studi Economico-aguridici della facoltà di Giurisprudenza dell'Università di Cagliari*, III, Part II.
- Gupta, S. S. and Huang, D. Y. (1976). On subset selection procedures for the entropy function associated with the binomial populations. *Sankhyā*, 38A, 153-173.
- Gupta, S. S. and Leu, L. Y. (1989). Selecting the fairest of  $k(\geq 2)$   $m$ -sided dice. Technical Report #89-16C, Department of Statistics, Purdue University, West Lafayette, Indiana.
- Gupta, S. S. and Wong, W. Y. (1975). Subset selection procedures for finite schemes in information theory. *Colloquia Mathematica Societatis János Bolyai*, 16, 279-291.
- Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* 58 13-30.
- Rizvi, M. H. Y., Alam, K. and Saxena, K. M. L. (1987). Selection procedure for multinomial populations with respect to diversity indices. *Contributions to the Theory and Application of Statistics*, A Volume in Honor of Herbert Solomon, ed. by Alan E. Gelfand, Academic Press, Inc. 485-509.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423, 623-656.
- Simpson, E. H. (1949). Measurement of diversity. *Nature*, 163, 688.

1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS			
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release, distribution unlimited			
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE						
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report #90-01C			5. MONITORING ORGANIZATION REPORT NUMBER(S)			
6a. NAME OF PERFORMING ORGANIZATION Purdue University		6b. OFFICE SYMBOL (if applicable)	7a. NAME OF MONITORING ORGANIZATION			
6c. ADDRESS (City, State, and ZIP Code) Department of Statistics West Lafayette, IN 37907			7b. ADDRESS (City, State, and ZIP Code)			
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research		8b. OFFICE SYMBOL (if applicable)	8. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-88-K-0170, NSF Grants DMS-8606964, DMS-8702620, DMS-8717799			
8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217-5000			10. SOURCE OF FUNDING NUMBERS			
			PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) SELECTING MULTINOMIAL POPULATIONS						
12. PERSONAL AUTHOR(S) Shanti S. Gupta and TaChen Liang						
13a. TYPE OF REPORT Technical		13b. TIME COVERED FROM TO		14. DATE OF REPORT (Year, Month, Day) January 1990		15. PAGE COUNT 12
16. SUPPLEMENTARY NOTATION						
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) fair multinomial population; entropy function; Gini-Simpson index; natural selection procedure; empirical Bayes selection procedure; asymptotically optimal; rate of convergence			
FIELD	GROUP	SUB-GROUP				
19. ABSTRACT (Continue on reverse if necessary and identify by block number) This paper deals with the problem of selecting fair multinomial populations compared with a standard. Two selection procedures are investigated: the natural selection procedure of Gupta and Leu (1989) and an empirical Bayes simultaneous selection procedure. It is proved that the natural selection procedure is a Bayes procedure relative to a symmetric Dirichlet prior distribution, and therefore is an admissible selection procedure. For the empirical Bayes simultaneous selection procedure, the associated asymptotic optimality is investigated. It is shown that the proposed empirical Bayes selection procedure is asymptotical optimal relative to a class of symmetric Dirichlet priors. The rate of convergence of the empirical Bayes selection procedure is shown to be of order $O(\exp(-\tau k + \ln k))$ for some positive constant $\tau$ , where $k$ is the number of populations involved in the selection problem.						
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified			
22a. NAME OF RESPONSIBLE INDIVIDUAL Shanti S. Gupta			22b. TELEPHONE (Include Area Code) (317) 494-6031		22c. OFFICE SYMBOL	